



# Communiquer avec les big data : vers une nouvelle naturalisation de l'attention des consommateurs ?

Jean-Sebastien Vayre

## ► To cite this version:

Jean-Sebastien Vayre. Communiquer avec les big data : vers une nouvelle naturalisation de l'attention des consommateurs ?. RT 11 – Sociologie de la consommation et des usages – Congrès de l'AFS 2015 “ La sociologie, une science contre nature ?, Jul 2015, Versailles Saint-Quentin, France. hal-01227396

**HAL Id: hal-01227396**

**<https://hal.science/hal-01227396>**

Submitted on 10 Nov 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Communiquer avec les big data : vers une nouvelle naturalisation de l'attention des consommateurs ?

Jean-Sébastien Vayre<sup>1</sup>

**Résumé :** La publicité est souvent considérée comme un vecteur de représentations stéréotypées qui participent à naturaliser l'attention des consommateurs. Une des promesses des technologies *big data* est de permettre aux offreurs de se défaire de ces formes d'essentialisation naïve en les amenant à rationaliser la conception de leurs communications. À travers une enquête de 36 mois reposant sur un ensemble de méthodes croisées nous montrons que les usages que les marchands font des technologies *big data* recouvrent des formes d'essentialisation naïve mais aussi et surtout d'essentialisation positive qui sont lourdes d'enjeux cognitifs et politiques.

**Mots-clés :** *big data*, marketing, communication, TIC, consommation, attention

**Abstract :** Advertising is often seen as a vector of stereotypical representations that contribute to naturalizing the attention of consumers. A promise of big data technologies is to enable suppliers to discard of these forms of naive essentialisation by leading them to rationalize the design of their communications. Through an investigation of 36 months based on a set of crossed methods we show that the uses that the merchants make of big data technologies cover forms of naive essentialisation but also and especially of positive essentialisation which are heavy cognitive and political issues.

**Key-words :** Big data, marketing, communication, ICT, consumption, attention

---

<sup>1</sup> [jsvayre@univ-tlse2.fr](mailto:jsvayre@univ-tlse2.fr), CERTOP - UMR 5044, Université de Toulouse II - Jean Jaurès.



Figure 1 : Illustration 1<sup>2</sup>

Durant les années cinquante, Moulinex lance une série de campagnes publicitaires particulièrement sexistes qui véhiculent des représentations stéréotypées des catégories « masculin » et « féminin ». De façon analogue aux jouets pour enfant (Garnier, 2012 ; Zegaï, 2010), ces campagnes publicitaires cristallisent des normes, des valeurs et des rapports de domination à travers les préjugés que leurs concepteurs se font de la différence sexuée.

D'un point de vue sociologique, ce type de communication pose des problèmes importants puisque, à la manière dont les équipements de marché peuvent participer à la performance des théories économiques (Callon, 1998), les campagnes « libère la femme » de Moulinex ont pu contribuer à celle des représentations stéréotypées que leurs concepteurs avaient des identités de genre. Autrement dit, comme le montrent Pascale Garnier (2012) et Mona Zegaï (2010) dans le cas des jouets pour enfant, ces représentations ont nécessairement joué un rôle non-

<sup>2</sup> Source : <http://www.live2times.com/1961-moulinex-libere-la-femme-e--8592/>

négligeable dans la construction des catégories « masculin »/« féminin », et donc, dans la stabilisation durable des rapports de domination qu'elles recouvrent.

Plus précisément, l'affiche présentée ci-dessus renvoie à une forme de naturalisation bien connue de la psychologie sociale : l'essentialisation que nous qualifions par la suite de naïve<sup>3</sup> (Rothbart et Taylor ; 1992). Les stéréotypes de genres véhiculés par la publicité Moulinex sont effectivement présentés comme naturels : le « Moulinex » est bien entendu destiné à « elle » qui, comme le connote sont tablier de cuisine, est par essence tournée vers la sphère domestique ; « lui », qui est naturellement tourné vers l'extérieur comme le symbolise son joli costume, peut alors profiter des « bons petits plats » qu'« elle » lui a préparé (cf. le « male breadwinner model » ; Pfau-Effinger, 2004). C'est pourquoi, comme nous le verrons par la suite, ce modèle de communication peut être qualifié d'heuristique dans le sens où il est conçu à partir de représentations qui peuvent faire l'objet d'un certain nombre de biais cognitifs (Tversky et Kahneman, 1974).

Aujourd'hui, il existerait une solution technologique permettant aux professionnels du marché (Cochoy et Dubuisson-Quellier, 2000) d'éviter ces formes d'essentialisation naïve. Cette solution, c'est le *big data*. À travers le traitement d'une immense quantité de données numériques, les technologies *big data* pourraient en effet garantir une véritable extraction de la réalité (i.e. : *reality mining* ; Pentland, 2014) et ainsi assurer le développement d'une connaissance objective telle qu'elle existerait empiriquement (Rouvroy et Berns, 2013). L'extrait d'entretien ci-dessous illustre comment les acteurs du *big data* marchand peuvent exprimer se positionnement :

« Historiquement, comment ils font pour faire de la publicité ? [...] Avant, ils faisaient fonctionner leurs préjugés, c'est-à-dire qu'ils disaient qu'il y a un concept qui s'appelle la ménagère de moins de 50 ans et elle va être intéressée par le robot ménager *Turbo 2000*. Sauf qu'en réalité, ce n'est pas du tout vrai ! Si vous regardez qui achète ces robots *Turbo 2000*, il y a à peu près entre 35 et 40 % d'hommes. [...] Quand on a des algorithmes qui sortent des résultats comme ça, on se demande si c'est l'algorithme qui déconne ou si c'est moi qui déconne ! Avec ce genre de méthode, vous allez apprendre de vos données [...].  
[Johann montre comment les techniques d'apprentissage artificiel (i.e. : *machine learning*) permettent l'apprentissage de concept<sup>4</sup> en référence au cas de *Netflix* et à partir de l'exemple du concept d'ultra violence]. Vous

---

<sup>3</sup> Nous précisons qu'il s'agit d'une forme d'essentialisation naïve dans le sens où elle est réalisée par les acteurs sociaux et de façon routinière. Nous verrons par la suite qu'il existe d'autres formes d'essentialisation qui sont plus positives dans le sens où elles renvoient à des postures technico-scientifiques particulières.

<sup>4</sup> La programmation logique inductive (PLI) permet par exemple de réaliser l'apprentissage artificiel de concept (cf. Vayre, 2014).

avez appris le concept d'ultra violence. C'est phénoménal du point de vue de l'analyse des données. De cette façon vous n'aurez pas besoin de valider des hypothèses abstraites. Au contraire, vous allez faire émerger des réalités sociologiques à partir des données en faisant un problème d'optimisation assez simple.

En jouant sur le *dataset* vous allez pouvoir apprendre des choses et vous allez ensuite pouvoir les réinjecter dans les modèles de données.

Le principe dans le *machine learning* c'est qu'on apprend les choses. [...] Pour chaque goût vous avez un modèle [...]. Vous regardez la variété du modèle et à quel moment votre performance plafonne. Il faut à peu près 500 paramètres en marketing. Donc vous apprenez un truc. Et, non seulement vous répondez au problème des philosophes qui voudraient qu'un concept est une catégorie d'analyse qui n'existe pas en soi, c'est une classe de construit [...]; non seulement ça existe mais la dimension c'est 500. C'est bluffant ; c'est très fort le *machine learning*. » (Johann, fondateur d'une start-up proposant des outils de personnalisation de la relation client, 2013).

Ainsi, comme le présente Johann, la force des *big data* couplées aux techniques d'apprentissage artificiel apparaît assez « bluffante » ; ou plutôt, intrigante puisqu'elle permettrait de déchiffrer la réalité sans pour autant faire appel à une quelconque forme de théorisation plus ou moins naïve<sup>5</sup>.

Afin de mieux comprendre les avantages et les limites des technologies *big data* telles qu'elles sont mobilisées dans le domaine du marketing, nous proposons donc de traiter le questionnement suivant : les technologies *big data* permettent-elles aux professionnels du marché de se défaire de toute forme d'essentialisation naïve vis-à-vis des consommateurs ? En favorisant la modélisation mathématique des comportements de ces derniers, ne recouvrent-elles pas une autre forme de naturalisation que l'on pourrait qualifier d'essentialisation positive ? Si oui, quels sont les enjeux cognitifs et politiques de cette « nouvelle » forme de naturalisation de l'attention des consommateurs ?

Afin de répondre à ce questionnement nous nous appuierons sur une enquête de trente-six mois reposant sur un ensemble de méthodes croisées de données qualitatives recueillies au cours de quatre observations (dont deux participantes réalisées pendant dix-huit mois auprès de deux *start-up* mobilisant les technologies *big data*) et au cours de présentations et d'entretiens réalisés par/auprès de plus de quarante professionnels du marché utilisant les technologies *big data*.

Ainsi, nous verrons que les *big data* qui intéressent les professionnels du marché constituent des données d'usage non-sollicitées qui peuvent renseigner les formes et les contenus de

---

<sup>5</sup> Notons que nous retrouvons ici l'idée de la fin de la théorie proclamée par Chris Anderson (2008).

l'attention des consommateurs. Nous montrons que ces dépôts d'attention favorisent le modèle de communication par percolation<sup>6</sup> et que ce dernier peut *in fine* renforcer les processus de performativité de la publicité (section 1). Ensuite, nous soutenons que si les *big data* doivent permettre de limiter les processus d'essentialisation naïve durant la conception publicitaire, les usages concrets qu'en font les professionnels du marché ne sont pas pour autant dépourvus de toute distorsion cognitive. Pour être analysés, les dépôts d'attention laissés par les consommateurs doivent faire l'objet d'une structuration qui peut recouvrir différentes formes de naturalisation par essentialisation naïve mais aussi et surtout par essentialisation positive (section 2). Finalement, nous soutenons que la sociologisation et la psychologisation des données, qui constituent des formes d'essentialisation positive, comportent des risques non-négligeables en termes d'organisation marchande (section 3). Nous concluons en soulignant que ces deux formes d'essentialisation positive renvoient à une posture d'ingénierie behavioriste des marchés qui pose d'importants problèmes sur les plans cognitifs et politiques.

### **1. Big data : dépôt d'attention et processus de percolation**

Les *big data* constituent un éventail de données numériques hétérogènes. D'une manière générale, il est possible de dégager deux grands types de données (Vayre, 2015). Le premier type est celui des données d'usages que les consommateurs et les travailleurs produisent quotidiennement et de façon plus ou moins volontaire à chaque fois qu'ils utilisent les Technologies de l'Information et de la Communication (TIC). Le second type est celui des données scientifiques, techniques et publiques dont font parties les *open data* et qui sont produites par les organismes publics (e.g. : *Météo France*, *INSEE*, les collectivités territoriales, etc.).

Du point de vue des professionnels du marché, les données d'usages sont particulièrement intéressantes parce qu'elles constituent des « dépôts d'attention » (Kessous, 2012) permettant de renseigner les comportements et les centres d'intérêts des consommateurs.

---

<sup>6</sup> Nous verrons par la suite que le modèle de communication par percolation (cf. Mallard, 2011) que recouvrent les technologies marchandes de type *big data* est bien différent du modèle de communication par heuristique qui caractérise la conception publicitaire traditionnelle. En effet, dans le premier cas, l'attachement entre l'univers de l'offre et celui de la demande est plus perméable et n'est pas intermédié par des représentations sociales plus ou moins stéréotypées mais par un processus de « re-présentation » (Latour, 1993) technico-scientifique de l'attention des consommateurs.

### ***1.1. Dépôts d'attention : des données issues de mesures passives renseignant les contenus et les formes de l'attention des consommateurs***

Comme le montre l'extrait d'entretien ci-dessous, les dépôts d'attention que composent les *big data* ont une particularité importante puisqu'ils constituent des données issues de mesures passives :

« [En référence aux dépôts d'attention que sont les *big data*] je pense que la tendance forte, c'est d'aller sur l'ensemble des médias vers des données de mesures passives [...] » (Morgane, directrice générale d'un institut d'enquête, 2013).

Que cela signifie-t-il que les *big data* constituent des données issues de mesures passives ? Cela signifie que ces données sont non-sollicitées dans le sens où elles ne sont pas construites selon une visée préalable par celui qui les recueille<sup>7</sup> et ne font pas l'objet d'une demande explicite à l'égard de celui qui les produit (Vayre, 2013). Ce point est central puisqu'il tend à renverser le rapport à l'enquête comme l'expose bien Morgane :

« On est des gens qui nous sommes construits sur l'idée de poser des questions, et que maintenant les questions sont secondaires, enfin sont secondes, plus exactement, dans le processus. C'est-à-dire qu'il y a un grand nombre de réponses et d'informations qui existent indépendamment des questions que nous pouvons nous poser. » (Morgane, directrice générale d'un institut d'enquête, 2013).

En outre, en référence aux travaux de Georg Simmel (1999), il est possible de classer les dépôts d'attention que composent une partie des *big data* en deux catégories : les données renseignant les contenus de l'attention et celles renseignant les formes de l'attention.

Plus précisément, en tant que consommateur et durant nos activités quotidiennes, nous mobilisons de façon plus ou moins intense les TIC. Aussi, durant ces activités, nous pouvons être amené à exprimer plus ou moins explicitement un certain nombre d'opinions à travers une interaction que nous pouvons partager avec un-e ami-e et/ou un-e collègue sur un site de réseautage de loisir et/ou professionnel ; un commentaire que nous exprimons sur un blog, un site d'actualité ou encore de e-commerce ; un *like* ou une note que nous accordons à autrui, à un service ou un produit ; une photo, une vidéo ou encore une musique que nous diffusons sur un site de partage culturel ; etc. Le web 2.0 est souvent qualifié de collaboratif. En ce sens, à

---

<sup>7</sup> Nous souhaitons ajouter que ce point n'est pas toujours vrai puisque certains professionnels du marché utilisant les *big data* décident du type de sonde qu'ils souhaitent mobiliser pour recueillir tel et tel type de données de telle et telle façon. Néanmoins, les dépôts d'attention sont généralement des données qui sont, encore aujourd'hui et pour la plupart, des données non-sollicitées.

chaque fois que nous participons à sa production, nous partageons (et rendons donc plus ou moins consciemment public) des idées, des opinions, des goûts, ou si l'on préfère, des représentations qui peuvent renseigner les contenus de l'attention que nous accordons à telle chose ou à telle personne (cf. Figure 2 : Illustration 2 ; Cardon, 2010). Ces informations peuvent être particulièrement intéressantes pour les professionnels du marché.

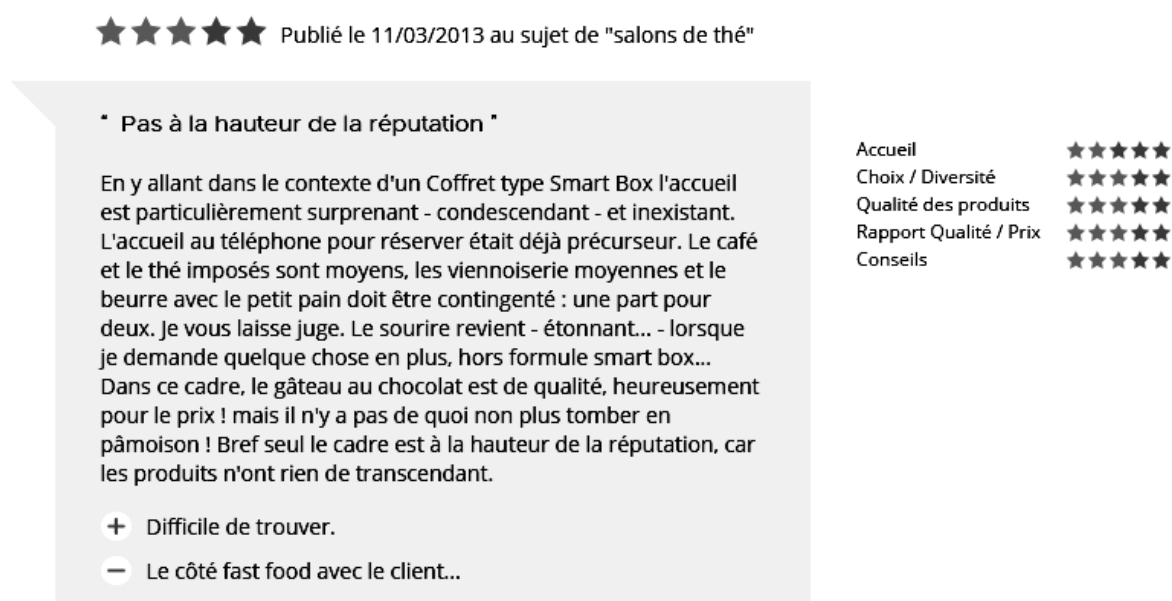


Figure 2 : Illustration 2

De plus, en tant que consommateur, à chaque fois que nous faisons usage des TIC, que ce soit pour partager ou non une représentation, nous produisons une très grande quantité de données qui peuvent être regroupées afin de retracer nos pratiques. Par exemple, lorsque nous visitons un site marchand, nos actions sont souvent enregistrées (cf. Figure 3 : Tableau 1). Plus précisément, l'ensemble des clics souris/tactiles, des actions-clavier et parfois des mouvements de souris que nous réalisons à l'intérieur du site sont capturés et généralement associés à des indicateurs de lieux (e.g. : coordonnées GPS), de temps (e.g. : *timestamp*), d'ancienneté (e.g. : *cookie* traçant le nombre de visites effectuées) voire aux données sociodémographiques lorsqu'elles sont disponibles. De même, les contenus des pages consultées font l'objet d'une analyse plus ou moins détaillée. De cette façon, en associant nos actions et les contenus des pages que nous avons consultées, les professionnels du marché peuvent saisir nos pratiques de navigation et ainsi comprendre les formes de l'attention que nous portons à l'égard de tel bien, service, commentaire, etc.



Exemple d'étiquette de variable	Valeur de la variable (et définition)
@fields.arg_ipClient	10.30.50.78 (adresse IP du visiteur)
@timestamp	2014-04-11T07:29:54.119Z (date et heure de l'évènement)
@fields.geoip.ip	109.10.125.56 (coordonnées GPS de lu visiteur)
@fields.arg_url	<a href="http://www.1001pharmacies.com/dodie-sucettes-0-6-mois-anatomique-silicone-message-maman-et-transparente-x2-p6651?utm_source=google-shopping&amp;utm_medium=comparateur-google&amp;utm_campaign=dodie-sucettes-0-6-mois-anatomique-silicone-message-maman-et-transparente-x2&amp;gclid=CNjfhs71170CFQolwwodELIAZQ">http://www.1001pharmacies.com/dodie-sucettes-0-6-mois-anatomique-silicone-message-maman-et-transparente-x2-p6651?utm_source=google-shopping&amp;utm_medium=comparateur-google&amp;utm_campaign=dodie-sucettes-0-6-mois-anatomique-silicone-message-maman-et-transparente-x2&amp;gclid=CNjfhs71170CFQolwwodELIAZQ</a> (URL de la page visitée)
@fields.arg_os	Windows (système d'exploitation mobilisé par le visiteur)
@fields.nb_visit	1 (nombre de visite du site réalisé par le visiteur)
@fields.page_type	["product"] (type de page consultée par le visiteur)
@fields.productView.price_c	2.96 (prix du produit consulté par le visiteur)
@fields.arg_pigReco	click (indique si la page consultée est issue d'un clic recommandation)
@fields.arg_reco_id	["6650"] (identité de la recommandation cliquée)
@fields.arg_reco_position	["0"] (position de la recommandation cliquée)
Etc.	

Figure 3 : Tableau 1

### ***1.2. Dépôts d'attention : processus de percolation et bouclage performatif***

Les dépôts d'attention laissés par les consommateurs durant les usages qu'ils font des TIC permettent ainsi de brouiller la frontière qui sépare l'univers de l'offre et celui de la demande (Vayre, 2015).

Du point de vue des professionnels du marché, les données d'usages favorisent ainsi le développement d'innovation communicationnelle reposant sur le modèle de la percolation (e.g. : le cas de Johann ; Mallard, 2011) dans le sens où elles permettent aux concepteurs d'accéder en temps réel aux formes et aux contenus de l'attention des consommateurs. Le modèle de la percolation s'oppose ainsi à celui de l'heuristique qui caractérise les modèles de conception publicitaire plus traditionnels et qui ne permettent d'accéder aux usages qu'aux travers des représentations sociales que partagent les concepteurs (e.g. : le cas de Moulinex ; cf. Figure 4 : Schéma 1).

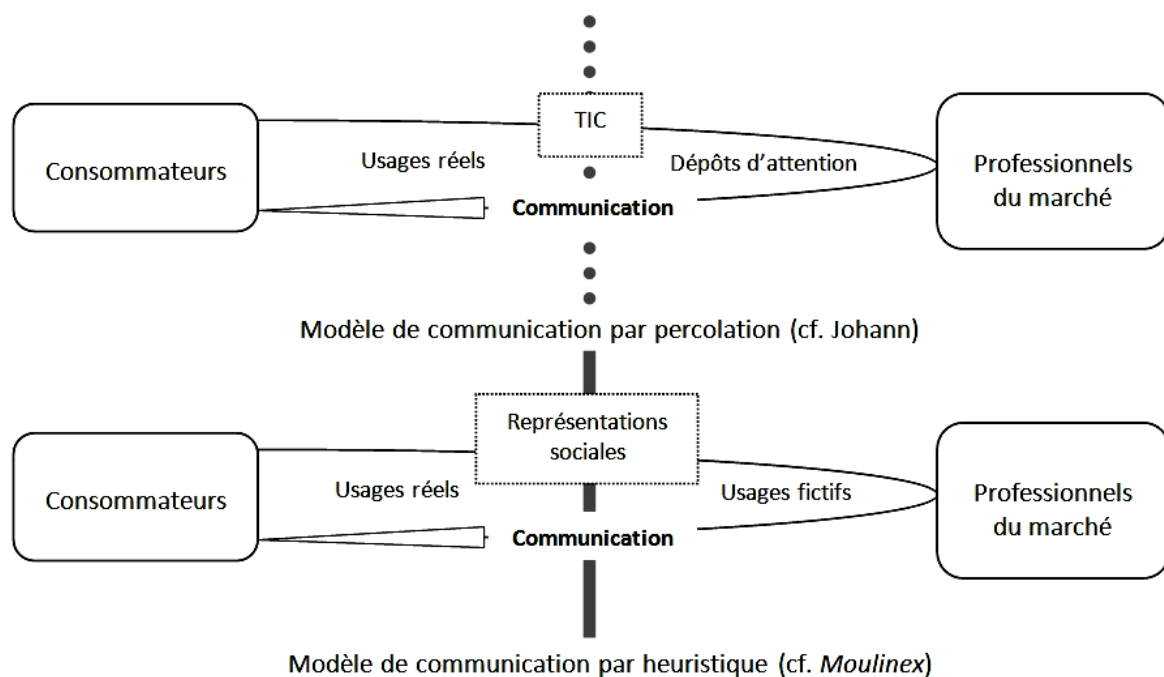


Figure 4 : Schéma 1

*A contrario* du modèle de communication par heuristique, celui dit par percolation permettrait ainsi aux spécialistes du marché de rester à l'écoute des consommateurs sans pour autant avoir à puiser dans le stock des représentations sociales plus ou moins stéréotypées qui sont à dispositions des concepteurs.

Autrement dit, compte tenu que les dépôts d'attention constituent une matière première produite en continu, les professionnels du marché qui en font usage auraient moins besoin de recourir à leurs intuitions (dont font partie les stéréotypes sociaux) pour se représenter les centres d'intérêts des consommateurs. Selon Johann, il leur suffirait en effet de trouver l'algorithme permettant de faire parler les données et, ensuite, d'apprendre en les écoutant discuter. Or, la linguistique pragmatique (Grice, 1975) et la psychologie cognitive (Chase et Simon, 1973) nous ont appris que toute forme d'inférence et d'apprentissage dépend des connaissances antérieures qui composent l'environnement cognitif de l'interlocuteur/apprenant.

Ainsi, pour Johann, mathématicien de formation, les données racontent que la représentation stéréotypée de la ménagère de moins de 50 ans ne réfère pas la réalité puisqu'entre 35 et 40 % d'acheteurs de robot *Turbo 2000* sont des hommes. Néanmoins, l'inférence que le mathématicien fait du récit des données nous semble poser problème. Retournons regarder l'affiche publicitaire de Moulinex. Qu'observons-nous ? « Lui » est en train de rentrer du travail avec un Moulinex comme cadeau pour « elle ». Les 35-40 % d'hommes acheteurs de

robot *Turbo 2000* ne pourraient-ils pas faire comme « lui » ? Si oui, les représentations statistiques de Johann corroboreraient parfaitement les représentations des concepteurs des campagnes publicitaires de Moulinex.

Ce renversement ne serait bien entendu pas très heureux. Toutefois, il ne serait pas dénué de sens sur le plan sociologique. Les données ne nous diraient pas que la représentation stéréotypée de la ménagère de moins de 50 ans est invalide mais plutôt qu'elle est une structure structurée (dans le sens où elle reflète l'incorporation par ses concepteurs de la division de sexe) fonctionnant encore aujourd'hui comme structure structurante (dans le sens où elle organise les pratiques et les perceptions des pratiques ; Bourdieu, 1979). L'attention portée par les hommes à l'égard du robot *Turbo 2000* pourrait donc révéler la profonde incorporation, par ces derniers, des stéréotypes de genre véhiculés par la publicité.

Notons que nous pourrions dire que, de toute façon, la compréhension approfondie de ce que raconte les données importe peu à Johann tend qu'elles lui permettent de faire de bonne prédiction. Il n'en reste pas moins que l'apprenant artificiel développé par Johann manque de connaissance écologique et sociohistorique sur les consommateurs afin de pouvoir évaluer la force prédictive de ce type de déterminants. Les *big data* sont encore trop peu nombreuses et pas assez diversifiées pour permettre à un apprenant artificiel d'objectiver de façon autonome le poids des processus de socialisation qui se jouent sur les marchés<sup>8</sup>.

En ce sens, et nous y reviendrons dans la section 3.1, si le modèle de communication par percolation que recouvrent les technologies marchandes *big data* peut effectivement conduire les professionnels du marché à considérer de façon plus empirique et moins stéréotypée l'attention des consommateurs, il peut également renforcer les effets performatifs d'autres publicités conçues selon le modèle de communication par heuristique en objectivant leur efficacité.

---

<sup>8</sup> Précisons que certaines méthodes d'apprentissage artificiel comme les inférences grammaticales (Higuera, 2005) ou les chaînes de Markov (Rabiner, 1989) pourraient techniquement permettre de modéliser ce type de phénomène. Néanmoins, les processus de socialisation qui se jouent sur les marchés, qu'ils soient primaires ou secondaires, débordent très largement le cadre des interactions numérisées et renvoient ainsi à une complexité psychologique et sociale que les *big data* ne permettent pas de mesurer. De plus, il s'agirait alors d'objectiver la construction de l'habitus (chez Pierre Bourdieu ; 1979), ou encore des habitus (chez Bernard Lahire ; 2006), ce qui impliquerait le stockage et le traitement d'une quantité de données très importante compte tenu qu'il s'agirait de se placer au niveau du parcours biographique de l'individu. À notre connaissance, l'écosystème matériel que constitue le système sociotechnique *big data* ne peut pas encore permettre ce type de traitement.

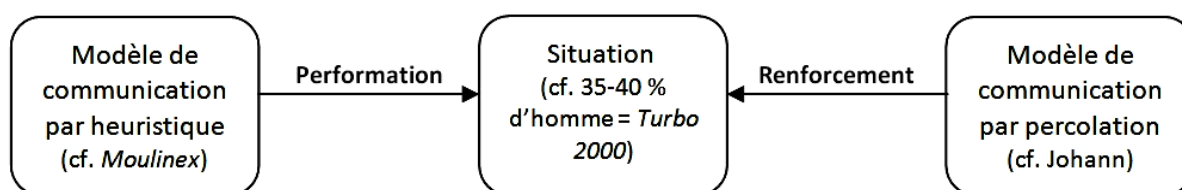


Figure 5 : Schéma 2

Ceci permet de souligner l'importance, sur le plan éthique, de ne pas utiliser les techniques d'apprentissage artificiel dans le seul but de pouvoir prédire efficacement la complexité de la dynamique sociale mais aussi de pouvoir la comprendre profondément.

## 2. Big data : biais de traçabilité, sociologisation et psychologisation des données

En outre, les usages que les professionnels du marché font des dépôts d'attention ne sont pas dépourvus de diverses formes d'essentialisation plus ou moins naïves qui ont des implications fortes sur les plans cognitifs et politiques.

Si les critiques positives (e.g. : Pentland ; 2014) et négatives (e.g. : Rouvroy et Berns, 2013) du *big data* mettent souvent en avant que ces dernières permettraient d'accéder aux phénomènes empiriques de façon directe et immanente<sup>9</sup>, nos observations empiriques mettent en avant que les *big data*, comme tout autre type de données, doivent intégrer une chaîne de traduction consistant à réduire et amplifier la réalité de façon à pouvoir la représenter (Latour, 1993). C'est pourquoi, les *big data* n'échappent pas au travail de structuration des données qui, opéré de façon plus ou moins automatique, recouvre un certain nombre de biais.

### 2.1. Biais d'apprentissage et de traçabilité

Du point de vu de l'usage que les professionnels du marché font des *big data*, il est possible de dégager deux grands types de biais : les biais d'apprentissage et les biais de traçabilité (cf. Figure 6 : Schéma 3).

À partir de nos investigations, nous pouvons dégager au moins deux catégories de biais d'apprentissage. La première est celle qui renvoie aux fameux compromis biais-variance qui caractérise toute technique d'apprentissage artificiel. Ce compromis peut être exposé à partir du dilemme suivant. Si le problème d'apprentissage comporte un trop grand nombre de variables, l'apprenant artificiel risque de ne pas trouver la solution du problème. Néanmoins, si le problème d'apprentissage ne comporte pas assez de variables, la solution trouvée par

<sup>9</sup> Un peu comme si les *big data* constituaient une sorte de papier-carbone de la réalité (Boullier, 2015).

l'apprenant risque d'être faussé (Vayre, 2014). Aussi, il est important de souligner qu'une machine ne peut pas réaliser une tâche d'apprentissage si ce dernier n'est pas biaisé. Cependant, un bon apprentissage artificiel ne doit pas être trop biaisé.

Le compromis biais-variance montre ainsi l'importance du travail de sélection de variable (i.e. : *featuring*) pour tout type d'apprentissage artificiel. Ce qui nous conduit à aborder la deuxième catégorie de biais d'apprentissage. Car, le processus de sélection de variables, qui est souvent associé à un travail de codage/recodage et/ou de calcul d'une ou plusieurs variable(s), peut être réalisé manuellement ou automatiquement. Et, lorsqu'il est effectué manuellement, ce processus peut avoir pour effet d'introduire un certain nombre de biais à travers les connaissances plus ou moins intuitives que les professionnels du marché ont vis-à-vis du problème d'apprentissage. C'est précisément en ce sens qu'une bonne partie des spécialistes du *big data* marchand que nous avons observés/interviewés soulignent l'importance des connaissances métiers et de l'intuition lorsqu'ils cherchent à paramétrer un problème d'apprentissage artificiel. Or, comme nous le verrons par la suite, les connaissances métiers et les intuitions des professionnels du marché peuvent parfois être empreintes de diverses formes d'essentialisation naïve. En outre, lorsque la sélection des variables est réalisée automatiquement, seul le niveau d'apport de chaque variable à la performance du modèle qui doit être appris est considéré. Mais, encore une fois, il est possible qu'un certain nombre de biais soient introduits à travers une ou plusieurs variables dites de confusion.

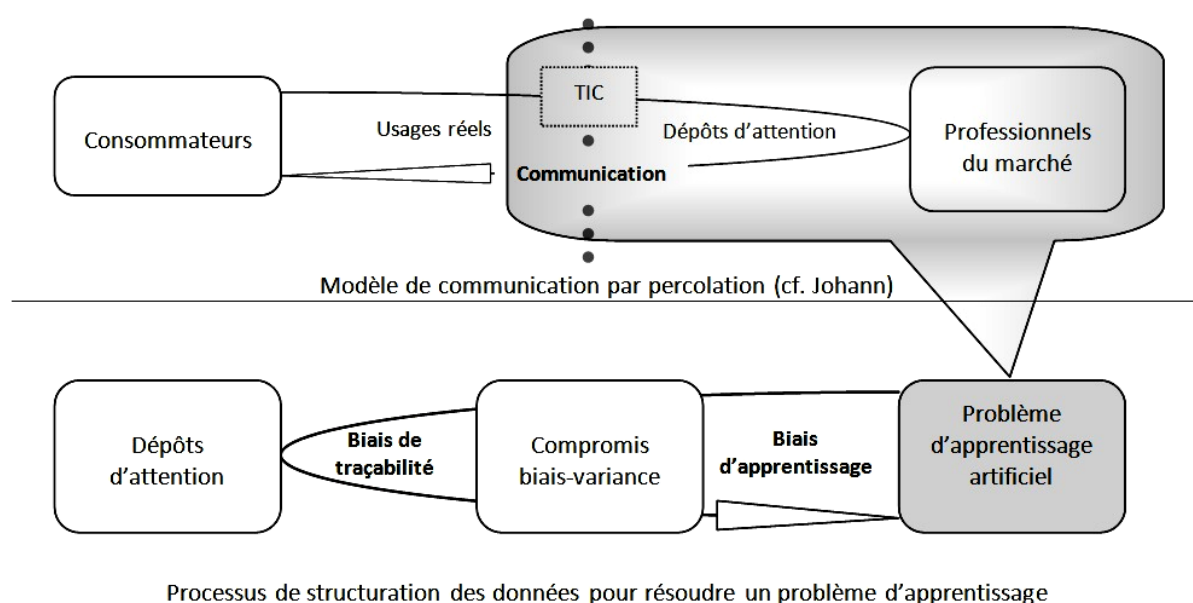


Figure 6 : Schéma 3

Ensuite, les biais de traçabilité peuvent être de formes très diverses. Ils reposent essentiellement sur le fait que les dépôts d'attention ne sont pas des données sollicitées ; et donc, ne font pas l'objet d'une méthode contrôlée cherchant à éviter les distorsions qui peuvent être introduites par la situation de recueil. Par exemple, dans le cas de l'analyse d'opinion qui intéresse beaucoup les professionnels du marché d'aujourd'hui, les avis/commentaires des consommateurs sont généralement compris d'une façon bipolaire (positif - neutre - négatif) et peuvent faire l'objet d'un certain nombre de biais (Boullier et Lohard, 2012). Pour illustration, les apprenants artificiels qui ont pour objet de classer les avis/commentaires ont actuellement encore beaucoup de mal à détecter les formes ironiques, l'humour, le second degré, le contrepied, etc. De plus, comme le souligne Yves Jeanneret dans le cas du guide touristique, les avis/commentaires déposés par les consommateurs ne sont pas simplement enregistrés par le site considéré : « [ce dernier] les façonne puisque son organisation documentaire fait programme et définit des styles de pratique » (2013 : 52). Ainsi, les avis/commentaires sont généralement produits dans des écologies sociomatérielles bien particulières qui ne sont quasiment jamais objectivées et qui confèrent pourtant une particularité importante aux données analysées<sup>10</sup>.

De même, dans le cas du e-commerce, les clics que les consommateurs réalisent sur les produits durant leur visite d'un site marchand sont souvent considérés comme des indicateurs d'intérêt. Or, les produits que les consommateurs consultent durant leur parcours de navigation ne sont pas nécessairement effectués par intérêt : dans certains cas de figure, ils pourraient plutôt indiquer une certaine désorientation (Amadieu *et al.*, 2008). De même, les durées de consultation des pages visitées sont également souvent considérées comme un indicateur d'intérêt par les professionnels du marché alors qu'elles peuvent révéler des difficultés de traitement, et surtout, sont généralement fortement corrélées aux niveaux d'expertise des consommateurs<sup>11</sup>. Soulignons alors que, dans ces deux cas de figure, associer le « clic produit » ou la « durée de consultation » à l'intérêt du consommateur constitue très clairement une forme d'essentialisation naïve des tâches que recouvrent les activités de recherche d'information.

---

<sup>10</sup> Il peut en effet exister différents effets de désirabilité sociale derrière les commentaires déposés par les consommateurs. De plus, les consommateurs actifs (i.e. : ceux qui tendent à partager facilement leurs opinions) comportent probablement des caractéristiques sociocognitives et sociodémographiques spécifiques qu'il conviendrait de mieux comprendre pour étudier rigoureusement la possible généralisation de leurs opinions (Cardon, 2010).

<sup>11</sup> Plus précisément, il ressort d'une analyse que nous avons effectuée de plus de 30 000 parcours de navigation réalisés sur des sites de e-commerce que plus les consommateurs connaissent le site visité et plus la durée moyenne de consultation des pages consultées tend à être réduite.

En d'autres termes, les dépôts d'attention laissés par les consommateurs durant les interactions qu'ils ont avec leur environnement numérique sont des indices extrêmement intéressants pour mieux comprendre les formes et les contenus de l'attention qu'ils portent à l'égard d'un site, d'une marque, d'un produit, d'un service, etc. Néanmoins, ces dépôts ne sont que des indices qui doivent faire l'objet d'une importante vigilance épistémologique afin de ne pas faire dire aux données ce qu'elles ne savent pas. En outre, comme nous l'avons vu précédemment, il est également important de souligner que toute forme d'apprentissage artificiel implique un travail de structuration qui, réalisé de façon plus ou moins automatique, peut recouvrir différents types de biais : les *big data* ne permettent pas d'accéder directement à une réalité qui existerait telle qu'elle ; comme toutes les autres données, elles s'insèrent au sein d'un processus de production de connaissance qui fait, par définition, l'objet d'une construction théorique et technique plus ou moins experte.

C'est pourquoi, si les technologies *big data* sont *a priori* une manière efficace d'éviter, ou plutôt, de limiter les biais de connaissance que recouvrent, par exemple, les formes d'essentialisation naïve, elles n'en sont pas non plus totalement dénuées.

## **2.2. Sociologisation des données**

À tout le moins, les technologies *big data* tels qu'elles sont utilisées dans les domaines de l'organisation et du marché renvoient à deux formes de naturalisation par essentialisation positive qui posent des problèmes sociocognitifs importants. La première est celle que nous qualifions de sociologisation des données. Elle prend sa forme la plus aboutie au travers de la notion de physique sociale telle qu'elle est réactualisée par Alex S. Pentland (2014).

Pour cet auteur, en constituant une manière de numériser la vie sociale telle qu'elle existe naturellement, les *big data* doivent pouvoir permettre sa mesure objective, et donc, sa modélisation mathématique. Nous retrouvons ainsi l'idéal des philosophes/sociologues les plus positivistes (e.g. : Comte, 2002 ; Durkheim, 2002 ; Saint-Simon, 2003) puisque cette modélisation mathématique de la société doit conduire l'observateur à découvrir les mécanismes de bases qui régulent les interactions humaines.

Cette sociologisation des données, que nous retrouvons presque explicitement chez Johann et de façon plus ou moins implicite chez l'ensemble des autres enquêtés, est une forme de naturalisation positive dans le sens où elle consiste à chercher à découvrir les structures surplombantes qui composent la « nature » du social. Comme dans le cas de la biologie, cette posture vise ainsi à dégager les déterminismes, ou si l'on préfère les lois naturelles, qui organisent, n'ont pas la physiologie individuelle, mais ce que l'on pourrait nommer la

physiologie sociale (Compte, 2002 ; Durkheim, 2002 ; Saint-Simon, 2003). Notons alors que l'intérêt des *big data* est, ici, de permettre de faire ressortir ces régularités sociales sans pour autant passer par des catégories d'analyse construites par l'observateur : les classes de groupe d'individu peuvent être apprises artificiellement en fonction des données disponibles les concernant.

Il n'en reste pas moins que la sociologisation des données que peut recouvrir les usages que les professionnels du marché font des *big data* est sous-tendue par un certain nombre de postulats qui renvoient à ce que nous pourrions qualifier, en référence aux auteurs classiques que sont John B. Watson (1924) ou encore Burrhus F. Skinner (1971)<sup>12</sup>, un behaviorisme sociologique. Elle repose en effet sur l'idée que les individus qui composent la société sont des agents sociaux qui s'activent essentiellement par routine et mimétisme. Autrement dit, ici, les agents sociaux incorporent les structures sociales de telle manière qu'ils constituent des sortes de programmes d'action qui sont essentiellement déterminés par le dehors : comme dans le cas des behavioristes classiques, ce qui se passe dans la tête de ces individus n'importe donc pas puisqu'ils agissent automatiquement en réponse aux stimuli de la société (*cf.* Figure 7 : Schéma 4). Ainsi, pour Alex S. Pentland (2014) les actions des individus sont déterminées socialement de telle sorte que la modélisation des structures sociales doit permettre de prédire les comportements humains à la manière dont on pourrait le faire pour des boules de billard. Par conséquent, nous percevons bien, ici, une théorie du déterminisme social qui n'est pas sans rappeler le « tout conditionnement » de John B. Watson (1924) ou encore de Burrhus F. Skinner (1971).

---

<sup>12</sup> Rappelons que pour ces deux figures classiques du behaviorisme, les états mentaux des individus ne sont pas directement observables et fonctionnent donc comme une boîte noire. De plus, ici, la créativité et l'invention n'existe pas puisque la conduite des comportements humains est exclusivement une histoire de conditionnement.



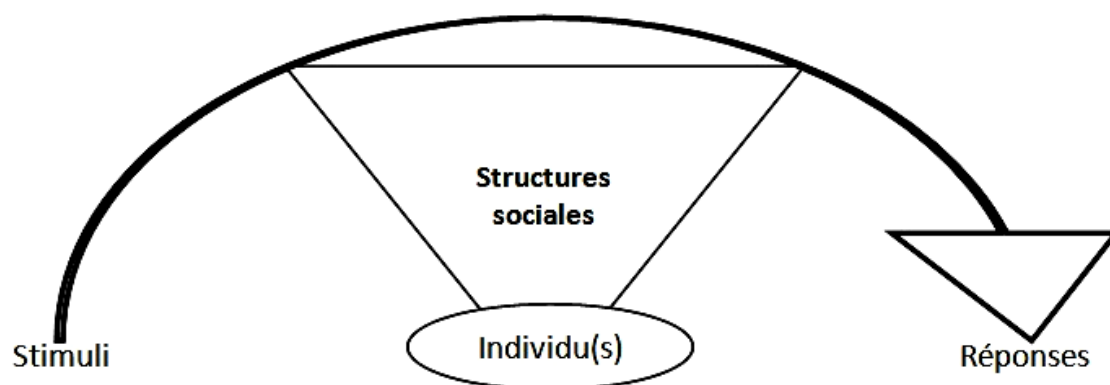


Figure 7 : Schéma 4

À la manière des premiers penseurs de la physique sociale (Compte, 2002 ; Durkheim, 2002 ; Saint-Simon, 2003) et comme nous le verrons dans la section 3.3, cette forme d'essentialisation positive qu'est la sociologisation des données conduit au développement d'une ingénierie behavioriste de type sociologique qui fait problème.

### 2.3. Psychologisation des données

Cependant, la sociologisation des données n'est pas la seule forme de naturalisation par essentialisation positive que nous pouvons retrouver plus ou moins explicitement chez les professionnels du marché qui font usages des technologies *big data*. Il existe en effet une autre forme d'essentialisation positive qui relève de ce que l'on pourrait nommer la psychologisation des données.

Cette fois-ci, la psychologisation des données renvoie à une posture plutôt de type évolutionniste dans le sens où elle consiste à rechercher des invariants dans les comportements de navigation des consommateurs de façon à pouvoir dégager des déterminismes universels associés au bon traitement ou non des informations. Aussi, bien que cette posture ne soit pas toujours explicite chez les spécialistes du *big data* marchand que nous avons observés/interviewés, elle n'en reste pas moins largement répandue. Par exemple et de manière implicite, nous retrouvons cette forme de psychologisation des données lorsqu'une campagne de marketing automatique est testée selon la méthode du *AB Testing* puisque celle-ci a généralement pour objectif d'évaluer les effets de la campagne considérée du point de vue de la manière dont elle impacte l'attention de l'ensemble des consommateurs.

D'une certaine façon, cette psychologisation des données constitue la symétrie inversée de la sociologisation des données. En effet, elle repose sur le postulat plus ou moins explicite que les individus sont déterminés, non plus par le haut (i.e. : par les structures sociales surplombantes), mais par le bas ; c'est-à-dire, par les ressources/contraintes cognitives que la cognition humaine a hérité de la phylogénèse (Chomski, 1986 ; Sperber, 1996). Ici, les individus ne sont plus des agents qui, par le biais de routines et de mimétismes, sont agis de l'extérieur, mais des êtres naturellement intelligents qui, par le biais de processus cognitifs infra-attentionnels, sont agis de l'intérieur.

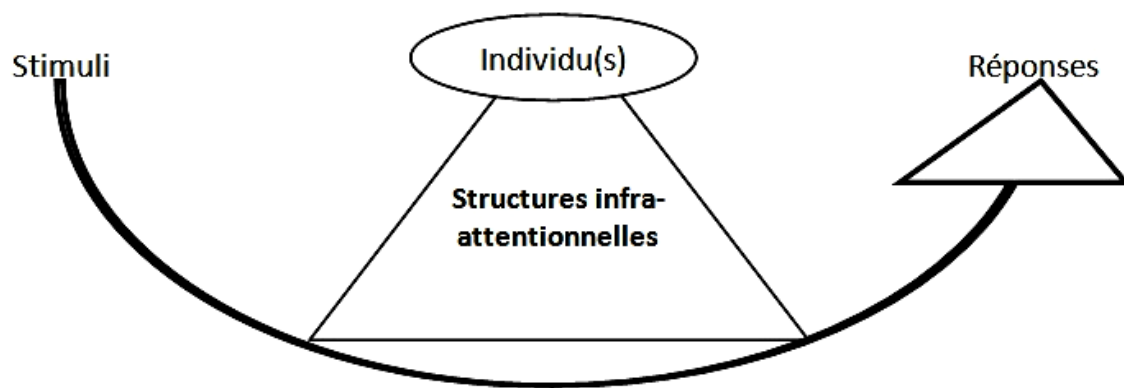


Figure 8 : Schéma 5

Comme dans le cas de la sociologisation des données, leur psychologisation peut être considérée, dans une certaine mesure, comme une forme de behaviorisme cognitiviste puisque, ici, il s'agit d'inférer l'universalité des structures infra-attentionnelles des cerveaux des consommateurs à partir de leurs comportements afin de pouvoir prédire la manière dont ils répondent aux stimuli environnementaux. De façon analogue à la sociologisation des données, nous verrons par la suite que leur psychologisation conduit au développement d'une ingénierie behavioriste de type psychologique qui pose également un certain nombre de problèmes sur les plans cognitifs et politiques.

### **3. Big data : une nouvelle naturalisation de la consommation ?**

Dans le domaine de l'organisation du marché, une des plus grands promesses du *big data* est de permettre la rationalisation des anticipations des décideurs. Rappelons en effet que, de manière traditionnelle en sciences humaines et sociales, ces dernières sont considérées comme des intuitions, des jugements (Knight, 1921), des actions non-logiques (Pareto, 1981), ou encore, des manifestations des « esprits animaux » (Keynes, 1972) qui se distinguent nettement de la rationalité substantive (Simon, 1976). En ce sens, les technologies *big data* doivent en quelque sorte permettre la rationalisation de ce que Gilbert de Terssac nomme le travail d'organisation (2002).

Comme dans le cas de la rationalisation de l'organisation du travail, celle du travail d'organisation pose bien entendu de très nombreux problèmes sociaux et humains. Néanmoins, dans le cas du marketing, il semble qu'un de ces avantages soit de conduire les professionnels du marché à développer des campagnes de communication publicitaire moins empreintes d'essentialisation naïve et plus attentives à l'attention des consommateurs.

#### ***3.1. Des enjeux cognitifs à ne pas négliger***

En effet, bien que les *big data* ne constituent pas un papier-carbone de la réalité (Boullier, 2015) et qu'elles puissent faire l'objet de divers biais (et notamment d'une certaine forme d'essentialisation naïve), elles n'en restent pas moins autant d'indices qui sont une source de connaissance considérable.

Les *big data* peuvent permettre aux professionnels du marché d'accéder à des connaissances empiriques particulièrement intéressantes compte tenu de la diversité et de la masse de dépôts d'attention qu'elles recouvrent. Aussi, ces connaissances peuvent faire l'objet d'un questionnement et d'une valorisation qui peut être considérée comme plus ou moins éthique du point de vu des consommateurs. La publicité renvoie en effet à deux grands types d'objectifs : documenter (i.e. : faire connaître un produit) et manipuler (i.e. : faire acheter un produit). Les professionnels du marché peuvent donc décider d'utiliser les *big data* dans le but de favoriser ou non l'un ou l'autre de ces objectifs.

Dans tout les cas, si les *big data* sont de formidables sources de connaissance qui peuvent effectivement permettre aux spécialistes du marché d'objectiver leurs savoirs et leurs pratiques, il ressort que, comme tout autre type de données, elles comportent un certain nombre de limites. Par exemple, elles ne sont pas encore assez nombreuses pour permettre à leur utilisateur d'accéder aux processus extrêmement complexes qui constituent, par exemple, la socialisation d'un individu (cf. note de bas de page 8). Les techniques plus traditionnelles que sont l'entretien, l'observation ou le questionnaire peuvent quant à elle relever ce défi à

condition qu'elles soient correctement conçues et analysées. Ces techniques de recueil de données constituent en effet des moyens d'accéder à une connaissance qui est actuellement très difficile à reconstruire à partir des *big data*<sup>13</sup>. De même, bien que les données comportementales puissent constituer des indices des états mentaux des consommateurs, elles ne permettent pas, par exemple, de repérer pragmatiquement les processus cognitifs impliqués dans les activités de recherche d'information comme peut le faire, par exemple, l'oculomètre, l'électroencéphalographie ou encore les entretiens d'auto-confrontation. En outre, les *big data* reposent généralement sur l'analyse de données indiquant des situations passées et/ou présentes. Or, il est clair que les représentations que les individus se font du futur jouent un rôle important dans la manière dont ils orientent leurs actions. Les individus ont des capacités de réflexivité : ils savent justifier et planifier leurs activités (Baars, 1997 ; Block, 1995). Et, les *big data* ne permettent que difficilement ou partiellement d'accéder à ce type de connaissance *a contrario*, par exemple, des entretiens collectifs (i.e. : *focus group*) qui sont, entre autre, de bons outils d'évaluation des intentions d'usage.

En synthèse, les *big data* sont des données qui, en tant que telles, comportent des avantages et des limites. En ce sens, elles doivent faire l'objet d'une vigilance épistémologique forte et/ou doivent être complétées/vérifiées à l'aide d'autres types de données pour pouvoir renseigner rigoureusement les formes et les contenus de l'attention des consommateurs. Notons néanmoins qu'il est également possible de les étudier pour elles-mêmes et de façon strictement pragmatique en faisant ce que Dominique Boullier appelle une sociologie des répliques (2015).

### **3.2. Les risques de bouclage performatif**

En outre, comme nous l'avons vu dans la section 1.2, si les *big data* doivent limiter les formes d'essentialisation naïve du point de vue du processus de conception publicitaire, elles peuvent par contre renforcer la force performative de la publicité traditionnelle.

---

<sup>13</sup> Notons que l'entretien, l'observation ou le questionnaire ont à leur tour de nombreuses limites que nous ne pouvons pas exposer ici. En outre, en référence aux travaux de Dominique Boullier (2015) qui voit dans les *big data* une troisième génération de sciences humaines et sociales, nos propos ne consistent pas à faire un plaidoyer pour les premières générations. Les *big data* permettent effectivement d'accéder à des formes de connaissance qui n'étaient pas accessibles jusqu'ici. En ce sens, la troisième génération de sciences humaines et sociales mise en avant par Dominique Boullier est bien entendu d'un intérêt considérable. Pour autant, nous pensons que, encore aujourd'hui, un nombre important de dimensions de la vie humaine et sociale ne peut pas être pris en compte par les technologies *big data* : par exemple, bien qu'il existe actuellement une diffusion toujours plus grande des objets connectés, un panel trop large des activités des consommateurs n'est pas numérisé et ne peut donc pas être mesuré et objectivé. En ce sens, il nous semble que, pour le moment et *a minima* dans certains cas de figure, une approche transgénérationnelle des sciences humaines et sociales constitue un compromis intéressant.

En effet, les *big data* telles qu'elles sont mobilisées par les professionnels du marché permettent le plus souvent de faire des analyses plutôt situées des consommateurs qui ne sont que rarement considérés d'un point de vue sociohistorique<sup>14</sup>. Aussi, comme nous l'avons vu dans la section 1.2, en tant qu'indices des formes et des contenus de l'attention des consommateurs, les *big data* peuvent finalement renseigner le pouvoir performatif de la publicité à travers l'examen des représentations et des pratiques des consommateurs : que signifie qu'entre 35 et 40 % d'hommes achètent le *robot Turbo 2000* ? Que les représentations stéréotypées des concepteurs des campagnes Moulinex sont fausses ou bien qu'elles ont plutôt fini par contribuer à l'institution des rôles de sexe ?

Du point de vu des professionnels du marché, les *big data* doivent permettre de mieux comprendre les consommateurs de façon à pouvoir trouver et utiliser les formes de communication qui sont les plus efficaces. Or, une communication efficace n'est pas nécessairement une communication pertinente dans le sens où elle répond aux besoins d'information de l'interlocuteur (Sperber et Wilson, 1986). Une communication efficace peut être une communication performative dans le sens où le locuteur peut faire advenir la situation en l'exprimant (Austin, 1962). Autrement dit, les *big data* pourraient parfaitement être considérées comme des indices montrant l'efficacité de l'institutionnalisation des règles, normes et valeurs véhiculées par la publicité. Comme nous l'avons dit dans la section 1.2, s'il y a entre 35 et 40 % d'hommes qui achètent le *robot Turbo 2000*, c'est peut-être parce que, ayant bien incorporés les rôles de sexe véhiculés par la publicité, ces derniers souhaitent l'offrir à une personne de sexe féminin. Auquel cas, les *big data* permettraient finalement de renforcer les pratiques sexistes par le biais, par exemple, de l'apprentissage d'un agent de recommandation qui pourrait alors proposer des *robots Turbo 2000* aux hommes visitant un site de e-commerce.

Ainsi, les *big data* doivent *a priori* conduire les spécialistes du marché à rationaliser leurs activités de conception publicitaire en limitant les formes de naturalisation par essentialisation naïve. Néanmoins ces derniers s'intéressant surtout au pouvoir prédictif des *big data*, celles-ci peuvent finalement conduire ces spécialistes à renforcer certaines pratiques de consommation qui pourraient alors être la résultante de l'incorporation de normes, règles et valeurs issues de représentations stéréotypées.

---

<sup>14</sup> À notre connaissance, les professionnels du marché n'analysent pas l'ensemble des données qu'ils possèdent sur les consommateurs mais les plus récentes seulement. Par exemple, dans le cas de la société *DataCrawler* qui développe un moteur de suggestion à destination des e-commerçants, seules les données recueillies durant les trois derniers mois sont prises en compte par les algorithmes de recommandation.

### ***3.3. Les risques de l'ingénierie behavioriste***

En outre, il apparaît que les deux formes de naturalisation par essentialisation positive que sont la sociologisation et la psychologisation des données recouvrent un certain nombre de risques politiques compte tenu que leur finalité consiste précisément à orienter les actions des consommateurs.

Bien entendu, un des premiers risques de l'ingénierie behavioriste que recouvrent les postures que sont la sociologisation et la psychologisation des données est qu'elle soit élaborée selon des intentions peu légitimes qui consisteraient à projeter la manipulation du consommateur au détriment de ses propres intérêts et dans le seul objectif de servir l'offreur. Néanmoins, même dans le cas où l'ingénierie behavioriste, qu'elle soit plutôt de type sociologique ou psychologique, apparaît recouvrir des intentions heureuses pour les consommateurs, elle n'en reste pas moins problématique sur le plan relationnel.

En effet, les deux approches que sont la sociologisation et la psychologisation des données sous-tendent l'idée que les individus agissent de façon automatique en réponse aux stimuli présents d'en l'environnement d'action. Selon ces deux perspectives, une manière d'orienter les comportements des consommateurs est alors de trouver les règles  $R_i$  qui déterminent les processus stimuli-réponse noté  $s_i \rightarrow r_i$  tel que  $R_i = s_i \rightarrow r_i$ . Dans le cas de la sociologisation des données, les règles  $R_i$  renvoient aux structures sociales ; dans le cas de la psychologisation des données, les règles  $R_i$  renvoient aux structures infra-attentionnelles. Dans les deux cas, les règles  $R_i$  doivent alors permettre de prédire les réponses  $r_i$  des consommateurs aux stimuli  $s_i$  et inversement ; soit,  $r_i = R_i(s_i)$  et  $s_i = R_i(r_i)$ . La sociologisation et la psychologisation des données doivent ainsi permettre aux professionnels du marché de concevoir des stimuli  $s_i$  de façon à conduire les consommateurs à adopter les réponses  $r_i$ .

Cette forme d'ingénierie behavioriste pose donc quelques problèmes sur les plans cognitifs et politiques. Prenons par exemple le cas du développement durable et imaginons qu'une équipe de professionnels du marché souhaite réaliser une sociologisation/psychologisation des données afin de conduire les consommateurs à adopter des pratiques plus soutenables. Les professionnels du marché vont alors chercher à développer un dispositif de communication en se basant sur le modèle  $s_i = R_i(r_i)$  afin de conduire les consommateurs à apprendre automatiquement de nouveaux comportements de consommation. Aussi, ce modèle pose problème dans le sens où, à la manière des behavioristes classiques, il ne tient pas compte de ce que la plupart des spécialistes de l'apprentissage savent bien depuis longtemps (Ausubel, 1968 ; Develay, 1995 ; Novak, 1977 ; Piaget, 1937) : l'apprenant est seul à construire son savoir et personne ne peut (ne doit ?) le faire pour lui. Autrement dit, en focalisant sur la

dimension automatique de l'action, la sociologisation et la psychologisation des données ne permettent pas de tenir compte des capacités de réflexivité des consommateurs. Pourtant, en tant qu'être humain vivant en société, ces derniers possèdent des compétences méta-cognitives leur permettant de comprendre une situation, de planifier leur action (Musial *et al.* 2012) et des compétences sociocognitives leur permettant de convoquer différents principes d'équivalence (i.e. : conventions) de manière à pouvoir évaluer et justifier ces actions (Boltanski et Thévenot, 1991). Et, c'est précisément à travers ces deux types de compétences que les consommateurs peuvent s'engager cognitivement dans les processus d'apprentissage qui sont au cœur de tout changement.

Par conséquent, chercher à orienter les consommateurs en jouant sur le caractère routinier et automatique de leurs comportements peut être une manière intéressante de les conduire à adopter de nouvelles pratiques de consommation à condition que les intentions de cette manipulation soit légitime (Thaler et Sunstein, 2008). Mais ce n'est pas toujours le cas. De plus, il est difficile de décider ce qui est ou non légitime pour autrui sans être certain de ne pas se tromper<sup>15</sup>. C'est précisément pourquoi la sociologisation et la psychologisation des données, en ne considérant pas les compétences métacognitives et sociocognitives des consommateurs, pose des problèmes de gouvernance non-négligeable : les consommateurs ont un rôle actif dans la construction de leur savoir et de leur pratique que la sociologisation et la psychologisation des données ne sait malheureusement pas bien considérer. Ne pas tenir compte de ce rôle constitue pourtant une manière d'instaurer une asymétrie d'engagement dans la construction des marchés qui comporte de nombreux risques sociaux et humains qu'il ne faut pas minimiser. Cette asymétrie pourrait en effet se manifester, de façon idéal-typique, par une distribution de l'attention marchande déséquilibrée : du côté de l'espace des offreurs se trouverait l'attention réflexive (ou, si l'on préfère, l'engagement cognitif) et du côté de l'espace des demandeurs l'attention automatique (ou, si l'on préfère, le désengagement cognitif).

---

<sup>15</sup> En ce sens, nous souhaitons ajouter la précision suivante. Comme le montrent Richard H. Thaler et Cass R. Sunstein (2008), une grande partie des comportements de consommation sont de l'ordre de l'automatisme. C'est d'ailleurs pourquoi les *nudges* sont des dispositifs permettant de conduire efficacement au changement des comportements des consommateurs. Néanmoins, ils sont incomplets sur les plans sociocognitif et politique puisqu'ils n'ont pas pour objectif d'engager cognitivement les consommateurs dans ces changements. En effet, l'engagement cognitif des consommateurs implique une certaine réflexivité qui n'est pas favorisée par les *nudges* et qui est pourtant garante, dans une certaine mesure, de leur autonomie. Ainsi, si l'on peut s'accorder pour dire que les *nudges* sont une manière d'atteindre efficacement le changement de comportement de consommation, ceux-ci ne sont pas suffisants puisqu'ils ne favorisent pas l'autonomie des choix des consommateurs. Or, comme le montre Albert O. Hirschman (1995), les choix des consommateurs sont aussi des manières de faire entendre leur voix.

## Conclusion

Dans le domaine du marketing, les technologies *big data* ont pour principale fonction de rationaliser le travail d'organisation du marché. En ce sens, elles constituent des outils de gestions qui posent de nombreux problèmes organisationnels. Néanmoins, un des avantages associés à leur usage marchand est, *a priori*, de tendre à limiter les processus d'essentialisation naïve dans la conception publicitaire. Toutefois, les usages concrets que les professionnels du marché font aujourd'hui des *big data* ne sont pas dépourvus d'un certain nombre de biais cognitifs.

La première forme de biais cognitifs que nous avons pu souligner est liée aux usages à visée essentiellement prédictive que les professionnels du marché font des *big data*. En ne cherchant pas à comprendre en profondeur ce que raconte les données, les professionnels du marché pourraient par exemple participer à renforcer, d'un point de vue sociohistorique, le pouvoir performatif de la publicité, et donc, les formes d'essentialisation naïve qui en sont sous-jacentes. En outre, le travail de structuration des données dont fait l'objet toute technique d'apprentissage artificiel peut également recouvrir divers biais cognitifs et, plus particulièrement, une certaine forme d'essentialisation naïve. Néanmoins, un des risques les plus importants associés aux usages marchands des technologies *big data* n'est pas celui de la naturalisation par essentialisation naïve mais plutôt celui de naturalisation par essentialisation positive. Comme nous l'avons vu, la sociologisation et la psychologisation des données peuvent en effet contribuer à instaurer progressivement une sorte d'ingénierie behavioriste marchande qui pourrait instituer une asymétrie d'engagement dans la production des marchés. Et, comme nous l'avons vu, cette asymétrie pose des problèmes relationnels importants.

## Références :

- AMADIEU Franck, Claude BASTIEN et André TRICOT (2008), « Les méthodes *on-line* 1 : analyse des parcours », dans Aline CHEVALIER et André TRICOT (dir.), *Ergonomie des documents électroniques*, Paris, Presses Universitaires de France, p. 251-270.
- ANDERSON Chris (2008), « The end of theory: the data deluge makes the scientific method obsolete », *Wired* (June 23).
- AUSTIN John L. (1962), *How to do things with words*, Oxford, Oxford University Press.
- AUSUBEL David P. (1968), *Educational Psychology A cognitive view*, New York, Holt Rinehart Winston.
- BAARS Bernard J. (1997), *In the Theater of Consciousness: the Workspace of the Mind*, New York, Oxford University Press.
- BLOCK Ned (1995), « On a confusion about a function of consciousness », *Behavioral and Brain Sciences*, 18, p. 227-287.
- BOLTANSKI Luc et Laurent THÉVENOT (1991), *De la justification : les économies de la grandeur*, Paris, Gallimard.



- BOULLIER Dominique et Audrey LOHARD (2012), *Opinion Mining et Sentiment Analysis : méthodes et outils*, Paris, Open Edition Press.
- BOULLIER Dominique (2015), « Vie et mort des sciences sociales avec le Big Data », *Socio*, (4), p. 19-37.
- BOURDIEU Pierre (1979), *La distinction. Critique sociale du jugement*, Paris, Édition de Minuit.
- CALLON Michel (1998), *The Laws of the Markets*, Oxford, Blackwell.
- CARDON Dominique (2010), *La démocratie Internet. Promesses et limites*, Paris, Seuil.
- CHASE William G. and Herbert A. SIMON (1973), « Perception in Chess », *Cognitive Psychology*, (4), p. 55-61.
- CHOMSKY Noam (1986), *Knowledge of language: its nature, origins and use*, New York, Praeger.
- COCHOY Franck et Sophie DUBUISSON-QUELLIER (2000), « Introduction. Les professionnels du marché, vers une sociologie du travail marchand », *Sociologie du travail*, 42 (3), p. 359-368.
- COMTE Auguste (2002), *Cours de philosophie positive 1<sup>ère</sup> et 2<sup>ème</sup> leçons* [En ligne]. [http://classiques.uqac.ca/classiques/Comte\\_auguste/cours\\_philo\\_positive/cours\\_philo\\_pos\\_1\\_2.pdf](http://classiques.uqac.ca/classiques/Comte_auguste/cours_philo_positive/cours_philo_pos_1_2.pdf). Page consultée le 1<sup>er</sup> juin 2015.
- DEVELAY Michel (1995), *Savoirs scolaires et didactiques des disciplines*, Paris, ESF.
- DURKHEIM Émile (2002), *Les règles de la méthode sociologique* [En ligne]. [http://classiques.uqac.ca/classiques/Durkheim\\_emile/regles\\_methode/durkheim\\_regles\\_methode.pdf](http://classiques.uqac.ca/classiques/Durkheim_emile/regles_methode/durkheim_regles_methode.pdf). Page consultée le 1<sup>er</sup> juin 2015.
- GARNIER Pascale (2012), « La culture matérielle enfantine : catégorisation et performativité des objets », *Strenæ* [En ligne]. <http://strenae.revues.org/761>. Page consultée le 1<sup>er</sup> juin 2015.
- GRICE Paul H. (1975), « Logique and conversation », in Peter COLE and Jerry L. MORGAN (dir.), *Syntax and Semantics 3: Speech Acts*, New-York, Academic Press, p. 41-58.
- HIGUERA (de la) Colin (2005), « A bibliographical study of grammatical inference », *Pattern Recognition*, (38), p. 1332-1348.
- HIRSCHMAN Albert O. (1995), *Défection et prise de parole*, Paris, Fayard.
- JEANNERET Yves (2013), « Faire trace : un dispositif de présentation du social », *Intellectica*, 1 (59), p. 41-64.
- KESSOUS Emmanuel (2012), *L'attention au monde. Sociologie des données personnelles à l'ère numérique*, Paris, Armand Colin.
- KEYNES John-Maynard (1972), *La théorie générale de l'emploi, de l'intérêt et de la monnaie*, Paris, Payot.
- KNIGHT Frank H. (1921), *Risk, Uncertainty and Profit*, Bostons, Houghton Mifflin.
- LAHIRE Bernard (2006), *L'homme pluriel : les ressorts de l'action*, Paris, Hachette Littératures.
- LATOUR Bruno (1993), « Le « pédofil » de Boa Vista – montage photo-philosophique », dans Bruno LATOUR (dir.), *Petites leçons de sociologie des sciences*, Paris, La Découverte, p. 171-225.
- MALLARD Alexandre (2011), « Explorer les usages : un enjeu renouvelé pour l'innovation des TIC », dans Julie DENOUEL et Fabien GRANJON (dir.), *Communiquer à l'ère numérique. Regards croisés sur la sociologie des usages*, Paris, Presses des Mines, p. 253-282.
- MUSIAL Manuel, Fabienne Pradère et André Tricot (2012), *Comment concevoir un enseignement ?*, Bruxelles, De Boeck.
- NOVAK Joseph D. (1977), *A theory of education*, New York, Cornell University Press.
- PARETO Vilfredo (1981), *Manuel d'économie politique*, Paris, Droz.

- PENTLAND Alex S. (2014), *Social Physics: How Ideas Turn Into Actions*, New York, Penguin Press.
- PFAU-EFFINGER Birgit (2004), « Socio-historical paths of the male breadwinner model - an explanation of cross-national differences », *The British Journal of Sociology*, 55 (3), p. 377-399.
- PIAGET Jean (1979), *Théories du langage, théories de l'apprentissage. Le débat entre Jean Piaget et Noam Chomski*, Paris, Seuil.
- RABINER Lawrence R. (1989), « A tutorial on hidden Markov models and selected applications in speech recognition », *Proceedings of the IEEE*, 77 (2), p. 257-286.
- ROUVROY Antoinette et Thomas BERNIS (2013), « Gouvernamentalité algorithmique et perspectives d'émancipation », *Réseaux*, 1 (177), p. 163-196.
- ROTHBART Myron and Marjorie TAYLOR (1992), « Category labels and social reality: Do we view social categories as natural kinds? », in SEMIN Gun and Klauss FIEDLER (dir.), *Language, interaction and social cognition*, London, Sage, p. 11-36.
- SAINT-SIMON Charles-Henri (2003), *La physiologie sociale. Œuvres choisies par Georges Gurvitch* [En ligne]. [http://classiques.uqac.ca/classiques/saint\\_simon\\_claude\\_henri/physiologie\\_sociale/st\\_simon\\_physio\\_sociale.pdf](http://classiques.uqac.ca/classiques/saint_simon_claude_henri/physiologie_sociale/st_simon_physio_sociale.pdf). Page consultée le 1<sup>er</sup> juin 2015.
- SIMMEL Georg (1999), *Études sur les formes de socialisation*, Paris, Presses Universitaires de France.
- SIMON Herbert A. (1976), « From substantive to procedural rationality », in Spiro J. LATSIS (dir.), *Method and Appraisal in Economics*, Cambridge, Cambridge University Press, p. 129- 148.
- SKINNER Burrhus F. (1971), *L'analyse expérimentale du comportement. Un essai théorique*, Bruxelles, Dessart et Mardaga.
- SPERBER Dan (1996), *La Contagion des idées. Théories naturalistes de la culture*, Paris, Éditions Odile Jacob.
- SPERBER Dan and Deirde WILSON (1986), *Relevance. Communication and Cognition*, Oxford, Blackwell.
- TERSSAC (de) Gilbert (2002), *Le travail : une aventure collective*, Toulouse, Octarès.
- THALER Richard H. and Cass R. SUNSTEIN (2008), *Nudge. Improving Decisions About Health, Wealth, and Happiness*, New York, Penguin Books.
- TVERSKY Amos and Daniel KAHNEMAN (1974), « Judgment under Uncertainty: Heuristics and Biases », *Science*, 185 (4157), p. 1124-1131.
- VAYRE Jean-Sébastien (2013), « Le big data et la relation client. Quand les traces numériques organisent l'échange marchand », *12ème J ournées Normandes de Recherche sur la Consommation « Société et consommation »*, [en ligne]. [https://www.researchgate.net/profile/Vayre\\_Jean-Sebastien/publication/271766651\\_Le\\_big\\_data\\_et\\_la\\_relation\\_client.\\_Quand\\_les\\_traces\\_numeriques\\_organisent\\_lchange\\_marchand/links/54d110ff0cf28959aa7a6305.pdf](https://www.researchgate.net/profile/Vayre_Jean-Sebastien/publication/271766651_Le_big_data_et_la_relation_client._Quand_les_traces_numeriques_organisent_lchange_marchand/links/54d110ff0cf28959aa7a6305.pdf). Page consultée le 1<sup>er</sup> juin 2015.
- VAYRE Jean-Sébastien (2014), « Comment coproduisons-nous notre environnement numérique marchand ? », *Ludovia « Les objets numérique création et consommation »*, [en ligne]. [https://www.researchgate.net/profile/Vayre\\_Jean-Sebastien/publication/271766632\\_Comment\\_coproduisons-nous\\_notre\\_environnement\\_numerique\\_marchand\\_/links/54d10fc00cf25ba0f040a40f.pdf](https://www.researchgate.net/profile/Vayre_Jean-Sebastien/publication/271766632_Comment_coproduisons-nous_notre_environnement_numerique_marchand_/links/54d10fc00cf25ba0f040a40f.pdf). Page consultée le 1<sup>er</sup> juin 2015.
- VAYRE Jean-Sébastien (2015), « Les Tableaux de Bord sur Données Massives, pour un nouveau management de l'innovation ? », *Innovations*, 2 (47), p. 101-121.
- WATSON John B. (1924), *Behaviorism*, New York, W.W. Norton.

ZEGAÏ Mona (2010), « La mise en scène de la différence des sexes dans les jouets et leurs espaces de commercialisation », *Cahiers du Genre*, 2 (49), p. 35-54.